Eye Movements are like Gestures in the Creation of Informal Algorithms

Robert Mackiewicz (rmackiew@swps.edu.pl)
SWPS University, Chodakowska 19/31. Warsaw, 03-815 Poland

Monica Bucciarelli (monica.bucciarelli@unito.it)
Università di Torino, Torino, 10124, Italy

Sangeet S. Khemlani (sangeet.s.khemlani.civ@us.navy.mil)
Navy Center for Applied Research in Artificial Intelligence, Naval Research Lab, Washington, DC 20375, USA

Philip N. Johnson-Laird (phil@princeton.edu)
Emeritus, Department of Psychology, Princeton University, Princeton, NJ 08540, USA
Department of Psychology, New York University, New York, NY 10003, USA

Abstract

People who have no experience with programming can create informal programs to rearrange the order of cars in trains. To find out whether they rely on kinematic mental simulations, the current studies examined participants' eye movements in two experiments in which participants performed various moves and rearrangements on a railway consisting of a main track running from left to right and a siding entered from and exited to the left track. In Experiment 1, they had to imagine different sorts of single moves of cars on the railway. The sequences of their fixations resembled iconic gestures: they tended to look at the starting location of the imagined move, and then at its final location. In Experiment 2, the task was to create descriptions of how to solve four sorts of rearrangements that differ in their Kolmogorov complexity. It predicted the time to find the correct solution and the relative number and duration of fixations recorded during the description of each move for rearrangements of different complexity. Participants were more likely to fixate on the symbols on the cars than anything else, and they fixated longer when the rearrangement was more difficult. They also tended to fixate regions of the tracks where a car's movement began or ended, as if they were imagining a car moving along the tracks. The results suggest that humans rely on a kinematic mental simulation when creating informal algorithms.

Keywords: abduction, eye movements, informal programming, kinematic mental models, simulations.

Introduction

An algorithm is a finite set of instructions whose execution can lead from a starting situation to an output (Knuth, 1997). When an algorithm is in a code that a computer can convert into commands to be executed, it is a computer program. Naive individuals who know nothing about programming can understand informal algorithms expressed in natural language, and even create them. Researchers have studied how programmers understand algorithms written in computer code (Aschwanden & Crosby, 2006) and what strategies they use in creating their own code (Davis & Zhu, 2022), but they have rarely studied how people with no experience in programming either understand or create informal algorithms. This neglect is surprising because these abilities help to identify the origins of recursive thinking, which is at the root of mathematics. The fundamental puzzle is: what mental processes and representations do people rely on to devise these informal algorithms or to deduce their consequences? In what follows, we describe a solution to this puzzle that relies on the kinematic simulation that reflects the temporal sequence of steps that an algorithm calls for.

When people describe a kinematic process, they often move their hands as if they were performing an imaginary action (Hadar & Krauss, 1999). They also imitate the dynamic properties of an object such as its rotation (Chu & Kita, 2011). Many gestures are spatial in nature, in that they connect inner thought with the outside world. The same is true for eye movements (Keogh & Pearson, 2011). When people describe a nearby object, they fixate it (Grant & Spivey, 2003). And when they imagine different transformations of the object, they tend to continue fixating it. For example, when they need to infer the answer to a question about a system of pulleys, they imagine animating one pulley at a time and fixate each one in its causal sequence (Hegarty, 1992). Eye movements may therefore play a similar role to gestures. The tendency to look at an object that one is thinking about is analogous to a deictic (pointing) gesture that picks out a particular referent (Shimojima & Katagiri, 2013). The tendency to follow visible or imaginary movements is similar to iconic gestures (Tversky, 2019).

Previous studies have examined how adults (Khemlani et al., 2013) and children (Bucciarelli et al., 2016; 2018; 2022) devised informal algorithms, and gestures played an important part in the process. Ten-year-old children made both deictic and iconic gestures when they described how to rearrange cars in a toy train that they were not allowed to touch or move. Adults do not seem to gesture as often as children, so the current study investigates eye movements instead. Thus, the aim of the present study is to test the predictions that eye movements and fixations should reflect a specific measure of the complexity of the algorithms (K-complexity) as well as the kinematic mental simulation that humans rely on to create the algorithms.

Eye-tracking provides valuable information about what individuals are aware of in the world, but it can be difficult to disentangle this relation in detail (Hyona, 2010). One way to do so is to combine eye-tracking with retrospective thoughts about how individuals tackled a problem (see e.g. Jarodzka et al., 2010). Another approach is to combine eye-tracking data with concurrent 'think aloud' protocols (Van Gog, Paas, & Van Merrienboer, 2005). Below, the article describes an environment that allows the simultaneous recording of verbal protocols and eye movements.

Previous studies with children (Bucciarelli at al., 2016; 2018; 2022) used a toy railway track that runs from left to right and has one siding entered from, and exited to, the left side of the track. In the current studies in a different laboratory, adult participants saw a similar railway environment on their computer screens (see Figure 1). The siding and the left track can both serve as temporary storage for cars. Each problem begins with a row of cars on the left track, and participants have to explain how to make a particular sort of rearrangement so that the cars arrive at the right track in the required order. Their description should apply to trains containing any number of cars. Only three basic moves are allowed in rearrangements:

S: move one or more cars from left track to the **S**iding,
L: move one or more cars from the siding to **L**eft track,
R: move one or more cars from left track to **R**ight track.

Once a car reaches the right track, it must stay there, because no move allows it to return to the left track. The participants in the experiments understood that cars cannot jump over each other, and that a car on the siding cannot go directly to the right track. There is an infinite number of potential rearrangements in the railway environment. Figure 1 presents the task of rearranging cars named with letters fedcba into the order fdbeca. This rearrangement is a *parity sort* as it groups cars from even number positions and odd number positions on the left track into the first and second half of the train on the right track.

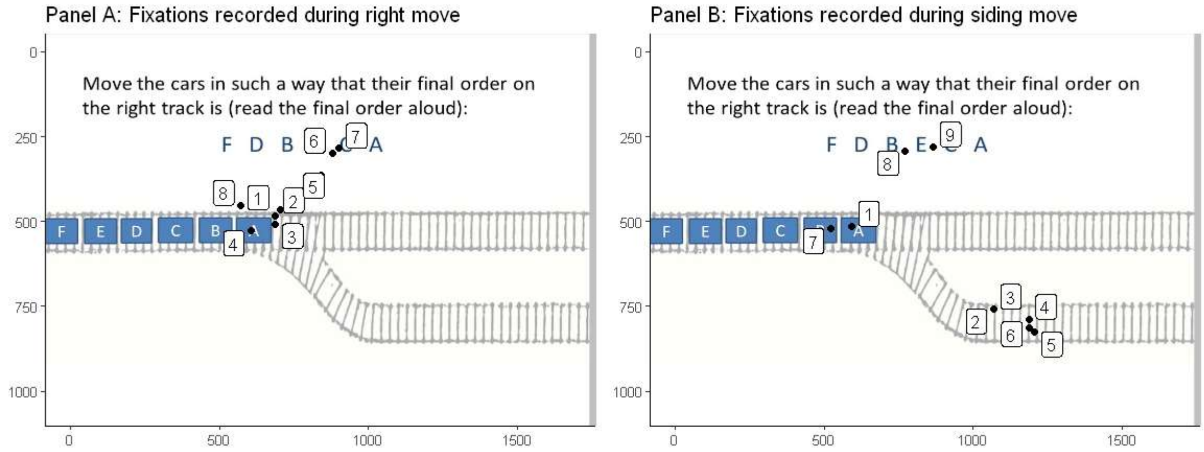

Figure 1: The computer display of the railway track in the initial situation presented to the participants in Experiment 2: tracks, the starting order, instructions and final order as indicated above the tracks (participants in Experiment 1 saw only the tracks). Numbered dots represent fixations recorded during a participant's description of the first two moves for the parity sort rearrangement in Experiment 2. Numbers refer to the order of fixations.

The sequence of required moves can be found by partial means-ends analysis, that is working backward from the required goal and adjusting the system to reduce the difference between the goal and the current state (see Newell, 1990). The process is a 'partial' means-ends analysis, because in the railway environment individuals envisage the moves for each of the positions in the target one at a time. As the first car at the head of the target order in parity sort is a, the first step is to get car a to the right track. The following diagram illustrates this move, which is also depicted in Panel A in Figure 1:

R1: fedcba [-] - ⇒ fedcb [-] a

In this notation, 'R1' denotes a move of one car to the right track. The square brackets represent the siding, the letters to the left of the brackets represent the contents of the left track, and the letters to the right of the brackets represent the contents of the right track, and so the hyphen represents zero cars. The next step is to move car b to the siding as the next car to enter the right track is c. The following diagrams represent these two steps, first of which is depicted in Panel B in Figure 1:

S1: fedcb [-] a ⇒ fedc [b] a R1: fedc [b] a ⇒ fed [b] ca

Some people may discover at this point that the further step is to repeat the same sequence S1R1 and then all cars from the siding have to be moved to the left track and then all cars from left have to be moved to the right track.

If people use partial means-ends analysis, then the difficulty of different rearrangements will depend on the difficulty in foreseeing the next move. This should be easier if someone grasps the general algorithm for a rearrangement. Such an algorithm for the parity sort can be expressed in computer code or in everyday language, for example in the following way:

1 Move one car to right track.
2 While there is more than one car on left track, move one car to siding, and move one car to right track.
3 Move one less than half the number of cars in the train to left track.
4 Move half the number of cars in the train to right track.

The algorithm for parity sort contains a loop S1R1 (line 2) and this loop is harder to discover than the loop (line 3 below) in the algorithm that reverses the order of cars:

1 Move one less than the cars to the siding.
2 Move one car to the right track.
3 While there are more than zero cars on the siding, move one car to the left track, move one car to the right track.

A simple measure of algorithm's difficulty is the Kolmogorov complexity (K-complexity for short; see Li & Vitányi, 1997). This is the length of the minimal program in a standard language for computing the output from a given input. A reasonable proxy that makes quite accurate predictions is the length of its description in English (see Johnson-Laird et al., 2022). The description above for parity sort contains 55 words whereas the description for reversal contains 40 words.

How do individuals discover general algorithms for rearrangement problems? The answer comes from the theory of mental models—the 'model' theory, for short. The main tenet of the theory is that people construct iconic mental models in their minds whose structure corresponds insofar as possible to the structure of what they represent. They represent relations, but not unnecessary details. In the case of the railway environment, models represent the positions of the cars on the tracks, but not the appearance of the cars. And they are kinematic, that is, they unfold in time to represent a temporal sequence of events. To figure out what steps are necessary to rearrange the order of a set of cars, individuals represent the initial arrangement, and they simulate each successive action.

Partial means-ends analysis, which underlies the creation of algorithms, implies that people simulate the moves of cars on the tracks move by move. Sometimes it is easier to foresee two or three moves in advance (as with reversal) and sometimes it is more difficult (as with parity sort). And it depends on the K-complexity. Eye fixations and movements should help people keep track of where the cars are at a given moment of mental simulation. This leads to the hypothesis that the description of moves for more complex algorithms should be accompanied by a larger number of fixations (Hypothesis 1). In addition, more complex rearrangements should require more attention, so that the mean duration of a single fixation should be longer (Hypothesis 2). Studies with children have shown that the deictic gestures used to point to a car and the iconic gestures used to show the movement of the car also reflect mental simulation. To pursue the parallel between gestures and eye movements, a plausible hypothesis is that individuals look at the cars whose moves they describe and follow their imagined moves with a sequence of fixations (Hypothesis 3).

The first experiment examined whether the railway environment is a suitable testbed for studying the relation between kinematic simulation and eye movements. Experiment 2 investigated how individuals deal with the discovery of an algorithm, and it tested the prediction that eye movements reveal kinematic mental simulations in the creation of informal algorithms.

Experiment 1

Method

Participants The experiment tested 22 students at SWPS University in Warsaw, Poland (8 men, 14 women, aged 19-23) in exchange for a course credit. Participants gave their informed consent before taking part in the experiment. This experiment and the following one were approved by the appropriate Ethical Committee.

Design and materials Participants acted as their own controls. They saw a computer monitor with a picture of the empty track, and their task was to imagine a car standing in one of the three main regions of the track and then making one of four moves from its initial location: S: from the left track to the Siding, L: from the siding to the Left track, R: from left track to the Right track, and a fourth move, which is not part of the standard set for rearrangements, i.e., from the right track to the left track. Participants had to imagine four instances of each sort of move, and each participant carried out the 16 trials in a different random order.

Procedure The experiment was conducted in the department's eye-movement laboratory. Each participant sat at a desk with a 26-inch computer monitor, with their eyes about 24 inches away from the monitor. Participants were asked to place their hands flat on the table and not to make any movements with their hands, head or shoulders, and were informed that their eye movements would be recorded by a device attached to the screen. They were informed about what moves are possible along the tracks and that their task is to imagine different sorts of moves. On each experimental trial, the participants heard a recorded description of the location of a car and then, after a short pause, a description of a move, which they had to imagine. They said, 'End', when they had done so. Here is an illustrative example:

> Imagine a car standing on the left track and look at its possible location. *(2 s silence followed).* Imagine that this car is moving to the siding.

When the participants said 'End', the experimenter pressed the button and the next trial began. Participants' eye movements were recorded using the SMI RED eye-tracking system at a sampling rate of 120 Hz.

Results

The eye tracker recorded the x- and y-coordinates of each fixation. These coordinates were classified into three regions: left track, right track, and siding. Following the recommendations of Holmqvist et al. (2011), each region had boundaries that were 1° viewing angle above and below the horizontal track and to the right or left of the diagonal part of the siding. All other fixations were classified as being outside the tracks. The results of four participants were excluded from statistical analysis, because the eye-tracking system was mis-calibrated for them.

Two time intervals were used to analyze the fixations. The first window was for fixations during the 2-second interval after the participants heard the description of a car's initial location. Participants made an overall mean of 5.18 fixations during this interval. They fixated the initial location of the car (M = 4.26 fixations) more often than any other area (M = 0.92: Wilcoxon test: z = 3.39, p < .001). Because they had been told to look at the part of the track where a move began, this result serves as an attention check. The second window was from the end of the verbal description of a move to the point at which the participants said 'End' to signal that they had finished imagining the move. The mean duration of this interval was 1.31s and the eye tracking system recorded 4.36 fixations. Participants fixated the correct destination of a move (M = 3.26) more often than any other area (M = 1.10: Wilcoxon test: z = 2.94, p = .002).

Discussion

The results suggest that eye fixations, like gestures, could be used to determine whether individuals use kinematic mental simulations when formulating rearrangement algorithms. If participants' eyes reflected imaginary moves, then they should fixate the correct start region and then the correct destination. This sequence occurred more often than not. Participants were instructed to look at the start regions, and therefore their fixations of destinations are more decisive.

Experiment 2

Method

Participants Twenty-six students form the same population as in Experiment 1 took part in the study on a voluntary basis (13 men, 13 women, aged 19-23). They gave their informed consent before participating in the experiment. They were told that the five who created the highest number of correct algorithms would receive a small financial reward (about $12).

Design and materials Participants acted as their own controls and carried out two instances of each of the four rearrangements: reversal, palindrome, parity sort, and faro shuffle. These rearrangements differ in K-complexity, as indicated by the length of their informal descriptions (number in brackets):

reversal (40) < palindrome (42) < parity sort (55) < faro shuffle (65)

Examples of reversal and parity sort have already been given in the Introduction. The palindrome algorithm changes the palindromic order abccba into pairs, as in aabbcc:

1 Move one less than half the cars to the siding.
2 Move two cars to the right track.
3 While there are more than zero cars on the left track, move one car to the left track, move two cars to the right track.

Faro shuffle interleaves cars from the second half into the first half, so that fedcba becomes fcebda:

1 Set the number of cars to be dynamically moved, n-of-s, to one less than half the cars.
2 Set decrement to one.
3 While n-of-s is more than zero, move one car to the right track, move n-of-s cars to the siding, move one car to the right track, move n-of-s cars to the left track, take decrement from n-of-s.
4 Move two cars to the right track.

Faro shuffle is the most complicated of the four algorithms; it uses a dynamic loop that changes the number of cars each time a loop is repeated, as shown by the length of the pseudocode above.

Each rearrangement was presented once with letters and once with numbers as labels on the cars, and the eight problems were presented to each participant in a different random order. The initial orders on the left track were fedcba for the letter trials and 654321 for the number trials (except for the palindrome, which had starting orders: abccba and 123321 to help participants grasp the structure of the rearrangement). The target order was displayed above the tracks (see Figure 1).

Procedure The procedure was the same as in Experiment 1, except that participants were told that their task was to describe as accurately as possible the sequence of moves required to move the cars from the left to the right track in a given arrangement, and that their voices were recorded using a dictaphone they held in their hands, a procedure that naturally prevented them from touching the screen and moving their hands. At the beginning of each trial, the participants read the target order aloud and soon after described the moves.

Results

The data from 3 of the 26 participants were excluded from the statistical analysis, as one participant discontinued the experiment, and two participants worked with a mis-calibrated eye-tracking system. So, the present results are for the remaining 23 participants. Two independent judges coded the recordings to make explicit the sequence of moves for each rearrangement and their accuracy. The judges agreed in their coding of accuracy on 94% of the trials (Cohen's κ = .93, p < .001). Discrepancies between them were resolved in discussion prior to the statistical analyses.

Behavioral data

There was no reliable difference between the accuracies of the descriptions of rearrangements for the two sorts of labels on the cars (77% correct descriptions for letters and 72% correct descriptions for numbers, Wilcoxon test, $z = 0.27$, $p > .75$) nor for the times for formulating correct descriptions (61.2 s for letters and 55.5 s for numbers; Wilcoxon test: $z = 1.27$, $p > .2$), and so the data from the two sorts of labeled cars were pooled for further analyses. Mean times for correct solutions were:

Reversals:	45.2s	Palindromes:	45.9s
Parity sorts:	71.7s	Faro shuffle:	70.5s

The participants' individual trends in these data corroborated the K-complexity prediction of difficulty (Page's $L = 444.0$, $z = 3.81$. $p < .001$). There were no reliable differences in accuracy for the four sorts of rearrangement (Page's $L = 542$, $z = 0.59$. $p > .5$).

Eye-tracking data

We analyzed the eye-tracking data for the correct descriptions of rearrangements for those participants who were correct at least once in each sort of algorithm. To ensure that the same amount of data was available from each participant, if someone was correct on two versions of the same algorithm (i.e., with letters and numbers), we randomly selected the solution to be analyzed. As a result, we had four verbal descriptions recorded from 16 participants. We matched these descriptions with the recording of the eye movements. We used the Audacity program to locate the beginning and end of each description of a move and matched these two points with the eye-tracking data. Table 1 shows the mean number of fixations recorded during the description of a single move and the mean duration of single fixations for each sort of rearrangement.

Table 1: Mean number of fixations and the mean duration of a single fixation (in milliseconds) recorded during the description of single moves in Experiment 2.

Rearrangement	Mean number of fixations	Mean fixation duration (in ms)
Reversal	7.7	228
Palindrome	8.6	228
Parity sort	9.8	239
Faro shuffle	10.5	238

The number of fixations during the description of a move in correct algorithms confirmed Hypothesis 1 about K-complexity (Page's $L = 428$, $p = .015$). Likewise, the mean duration of fixations during the descriptions of each sort of rearrangement corroborated Hypothesis 2 (Page's $L = 420$, $p < .05$).

Participants described a total of 693 moves, and we matched each move with a scan path, i.e., the accompanying series of fixations. Each fixation within a move description was categorized into one of five regions: left track, right track, siding, target order, and the instruction (see Figure 1), as in Experiment 1. There were 12 move descriptions where all fixations fell outside the three regions of the tracks. For the remaining 681 matches of move descriptions and fixations, participants looked at the left track, the right track or the siding at least once. In Experiment 1, we showed that participants looked at the tracks in the order predicted by the sort of move: first the origin of the move and then the destination. We used the same criterion here, except that we treated fixations as evidence for kinematic simulations if they followed the predicted order at each phase of the move description. We illustrate our approach with two sets of fixations, shown in Figure 1, in relation to the first two moves of parity sort. Panel A shows the description of the right move, but all fixations are on the left track. This type of scan path does not confirm that the eyes followed an imaginary move. The fixations in Panel B were recorded during the description of the siding move. The participant first fixated the left track and then moved his eyes to the siding. This sequence of fixations can be taken as evidence of a mental simulation, as it corresponds to a siding move. The scan paths corresponding to a move were recorded in 36% of the descriptions of right moves, 24% of the descriptions of siding moves and 20% of the descriptions of left moves. There were only 3 out of the 16 participants whose eye movements followed imaginary moves for half of the descriptions or more. This result is not surprising as in 56% of all matchings, participants only looked at the area with the squares representing cars: the left track and the final order above the tracks. In the remaining 44% of matchings, however, the participants looked on the right track or siding at least once. These glances were directed according to the sort of move. When describing moves to the right track, participants looked at the right track in 49% of moves descriptions, while they looked at the siding significantly less often (25% of moves; Wilcoxon test, $z = 2.56$, $p = .01$). Likewise, there were more descriptions of moves to or from the siding accompanied by fixation on the siding than fixation on the right track (23% vs 5%, Wilcoxon test; $z = 3.47$, $p < .001$).

Discussion

Experiment 2 showed that there was an increasing trend in the participants' numbers of fixations per move over the increasing complexity of four sorts of algorithm (Hypothesis 1). Likewise, there was an increasing trend in mean duration of a fixation with increasing complexity of the four sorts of algorithm (Hypothesis 2). Detailed examination of the sequence of fixations recorded during the description of a single move did not fully confirm the prediction that the eyes follow imaginary moves along the tracks (Hypothesis 3). In more than half of the descriptions of the moves, participants looked only at the left track and the final order. Consistent with the hypothesis that when moving their eyes along the tracks, participants should do so in a similar way to iconic gestures, they moved their eyes to the areas relevant to the described move.

General Discussion

The participants in our study described how they could rearrange the order of the cars of a train on a railway track using the siding, which can be used as a temporary storage. But they did not move any actual cars. The railway environment was displayed on a computer screen, and the system recorded the participants' descriptions of their solutions and their eye movements.

The aim of Experiment 1 was to test whether the railway environment is a suitable testbed for investigating the relationship between eye movements and mental simulations. When participants were asked to imagine single moves from one part of the railway to another, they tended to fixate the starting point and then the end point of the move. This result is consistent with the model theory and replicates the finding that when people imagine a scene, they move their eyes as if they were looking at the scene in reality (Johansson et al., 2012).

In Experiment 2, participants described the steps that lead to four sorts of rearrangements. The time they took to list the correct steps depended on the difficulty of the rearrangement algorithm. The new discovery was that this held also for how they moved their eyes during each single step: the Kolmogorov complexity (K-complexity) of an algorithm predicted the number of fixations they made during their description of the rearrangement (Hypothesis 1). K-complexity also predicted the mean duration of a single fixation (Hypothesis 2), which is typically considered an indicator of attentional demand (Rosch & Vogel-Walcutt, 2013). During the descriptions of the rearrangements, participants had a very strong tendency to keep their eyes on the letters or numbers representing cars, and only sometimes did they move their eyes elsewhere. However, when they moved their eyes along the track from one fixation to the next, they did so in a way that was consistent with the move they described (Hypothesis 3).

Is it possible to explain the present results without assuming that adults perform kinematic mental simulations? One explanation would be that they rely instead on a 'propositional' representation of the railway that has a grammatical structure analogous to a well-formed formula in a standard logic. The mental process would update a propositional representation, such as:

$$((\text{e})_{\text{left track}}\ (\text{f})_{\text{right track}}\ (\text{d c b a})_{\text{siding}})_{\text{railway}}$$

to:

$$((\ )_{\text{left track}}\ (\text{e f})_{\text{right track}}\ (\text{d c b a})_{\text{siding}})_{\text{railway}}$$

The three labeled brackets demarcate the three parts of the track, and their overall structure is equivalent to a phrase-structure tree, though the order of the terms within the brackets retains an iconic relation to the order of the cars on the relevant part of the track. The three sorts of move call for transformational rules to pass from one such tree to another (cf. Chomsky, 1957). This process may put more of a load on working memory than a kinematic simulation, but it is consistent with studies showing that the temporal order of fixations when events are imagined differs from the order when they are actually perceived (Gurtner, Hartmann, & Mast, 2021) and with studies showing that when participants recall a scene, they usually move their eyes to reflect the corresponding locations of objects in that scene (Martarelli & Mast, 2013). Yet, this interpretation fails to explain why eye movements are similar to the actual moves of the cars. It is harder to imagine the next move when the general algorithm for solving the rearrangement is more complex. And participants needed more time, and therefore fixated longer when they described moves in more complex rearrangements.

There was a clear difference between eye movements when participants were asked to imagine moves (Experiment 1) and when they described the moves while solving rearrangement problems (Experiments 2). Mental models underlie visual images, and visual images can even hinder reasoning (Knauff, 2009). Likewise, kinematic mental simulations can be more abstract than simulated actions. One need not take eight hours to imagine flying from New York to Paris and imagine the Eiffel tower (see Hesslow, 2012). The same process may have occurred in the present studies. Once participants knew how the cars moved to some location, they did not have to simulate their route in detail. At least they did not have to do that for every move. If they inspected the contents of their current state of their kinematic simulation, they knew when the cars were standing at the moment. It was enough to just look at the cars' locations every now and then or slightly move the eyes into the direction of a move.

Unlike other studies on eye movements, the present study did not measure them in short time windows, but recorded a whole series of eye movements during trials that could last almost 4 minutes. It is quite likely that there was some discrepancy between the timing of the verbal descriptions and the mental simulation of a move. There is no way to avoid the possibility of mismatch between the timing of simulation and the timing of naming the move in this type of study. Another difficulty in establishing the relations between kinematic simulations and eye movements lies in the fact that verbal descriptions of moves and eye fixations do not have to be strictly coordinated (Griffin & Bock, 2000). On the one hand, it is possible that the participants were thinking of a subsequent move when describing the current move. This ability, if it exists, could have led them to fixate the cars to be moved next. On the other hand, it is also possible that when people have to describe algorithms in their own words, the referents of those words could be a strong source of fixations. In general, the time interval between fixation and speech onset is not the same in all cases; it can vary between genders and depends on experience with the task (Law, Pellegrino, & Hunt, 1993).

The creation of an algorithm is a high-level skill — a way of thinking that goes beyond most thinking in daily life. The model theory and its computer implementation (see Khemlani et al., 2013) offer an explanation that has been proven in repeatable experiments. The central assumption is that kinematic simulations simulate the steps necessary to create informal algorithms.

Acknowledgments

The research was funded in part by the Polish National Science Center [Grant 2014/14/M/HS6/00916] (to RM) and by Fondazione Compagnia di San Paolo e Cassa Deposito e Prestito [Grant AI-LEAP N. D13C23001280007] (to MB).

References

Aschwanden, C., & Crosby, M. (2006). Code scanning patterns in program comprehension. In *Proceedings of the 39th Hawaii International Conference on System Sciences.*

Bucciarelli, M., Mackiewicz, R., Khemlani, S., & Johnson-Laird, P. N. (2016). Children's creation of algorithms: simulations and gestures, *Journal of Cognitive Psychology, 28*, 297-318.

Bucciarelli, M., Mackiewicz, R., Khemlani, S., Johnson-Laird, P.N. (2018). Simulation in children's conscious recursive reasoning. *Memory & Cognition, 46*, 1302–1314.

Bucciarelli, M., Mackiewicz, R., Khemlani, S. S., & Johnson-Laird, P. N. (2022). The causes of difficulty in children's creation of informal programs. *International Journal of Child-Computer Interaction, 31*, 100443.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*, 102–116.

Davis, D. K., & Zhu, F. (2022). Analysis of software developers' coding behavior: A survey of visualization analysis techniques using eye trackers. *Computers in Human Behavior Reports, 7*, 100213.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*, 462-466.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Gurtner, L. M., Hartmann, M., & Mast, F. W. (2021). Eye movements during visual imagery and perception show spatial correspondence but have unique temporal signatures. *Cognition, 210*, 104597.

Hadar, U. & Krauss, R.K. (1999). Iconic gestures: the grammatical categories of lexical aliates. *Journal of Neurolinguistics 12*, 1-12.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1084-1102.

Hesslow G. (2012). The current status of the simulation theory of cognition. *Brain Research, 1428*, 71–79.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford: Oxford University Press.

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172-176.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20(2),* 146-154.

Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology. Human Perception and Performance, 38*, 1289–1314.

Johnson-Laird, P. N., Bucciarelli, M., Mackiewicz, R., & Khemlani, S.S. (2022). Recursion in programs, thought, and language. *Psychonomic Bulletin & Review, 29(2)*, 430-454.

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *Plos one, 6*, e29221.

Khemlani, S.S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P.N. (2013). Kinematic mental simulations in abduction and deduction. *Proceedings of the National Academy of Science of the United States of America, 110*, 16766–16771.

Knauff, M. (2009). A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. *Spatial Cognition and Computation, 9(2)*, 109–137.

Knuth, D. (1997). *The art of computer programming. Volume 1: Fundamental algorithms. Third Edition*. Reading, Massachusetts: Addison-Wesley.

Law, D., Pellegrino, J., & Hunt, E. (1993). Comparing the tortoise and the hare: Gender differences and experience in dynamic spatial reasoning tasks. *Psychological Science*, 4, 35-40.

Li, M., & Vitányi, P. (1997). *An Introduction to Kolmogorov complexity and Its Applications (2nd ed.)*. Springer.

Martarelli, C. S., & Mast, F. W. (2013). Eye movements during long-term pictorial recall. *Psychological Research, 77(3)*, 303-309.

Newell, A. (1990). *Unified theories of cognition.* Harvard University Press.

Rosch, J.L. & Vogel-Walcutt, J.J. (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology & Work, 15,* 313–327.

Shimojima, A., & Katagiri, Y. (2013). An eye-tracking study of exploitations of spatial constraints in diagrammatic reasoning. *Cognitive Science, 37,* 211-254.

Tversky, B. (2019). *Mind in Motion: How Action Shapes Thought*. London: Hachette.

Van Gog, T., Paas, F., & Van Merrienboer, J. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.